TITLE
PLANT BRANCHED-CHAIN AMINO ACID BIOSYNTHETIC ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/063,423, filed October 28, 1997.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding enzymes involved in biosynthesis and utilization of branched chain amino acids in plants and seeds.

BACKGROUND OF THE INVENTION

Many vertebrates, including man, lack the ability to manufacture a number of amino acids and therefore require these amino acids preformed in their diet. These are called essential amino acids. Plants are able to synthesize all twenty amino acids and serve as the ultimate source of the essential amino acids for humans and animals. Thus, the ability to manipulate the production and accumulation of the essential amino acids in plants would be of considerable importance and value. Furthermore, the inability of animals to synthesize these amino acids provides a useful distinction between animal and plant cellular metabolism. This can be exploited for the discovery of herbicidal chemical compounds that target enzymes in the plant biosynthetic pathways of the essential amino acids and thus have low toxicity to animals.

The branched-chain amino acids leucine, isoleucine and valine are three of the essential amino acids. Biosynthesis of these amino acids proceeds, in part, via the common enzymes acetolactate synthase, acetohydroxyacid reductoisomerase, dihydroxyacid dehydratase and branched chain amino acid aminotransferase, and in part via enzymes specific for one of the amino acids, threonine dehydratase (isoleucine), and α-isopropylmalate synthase, 3-isopropylmalate dehydratase and β-isopropylmalate dehydrogenase (leucine). Regulation of the biosynthesis of each member of this family in plants is interconnected (see Figure 1), but understanding of the control is poor.

Few of the genes encoding enzymes that regulate this pathway in plants, especially corn, soybeans, rice and wheat, have been isolated and sequenced. For example, no plant genes have yet been reported for dihydroxyacid dehydratase, branched chain amino acid aminotransferase or 3-isopropylmalate dehydratase. Accordingly, the availability of nucleic acid sequences encoding all or a portion of these enzymes would facilitate studies to better understand the cellular control of the pathway, provide genetic tools for the manipulation of the pathway and provide a means to evaluate chemical compounds for their ability to inhibit the activity of these plant enzymes.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding plant enzymes involved in biosynthesis and utilization of branched-chain amino acids.

Specifically, this invention concerns an isolated nucleic acid fragment encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase. In addition, this invention relates to a nucleic acid fragment that is complementary to the

5   nucleic acid fragment encoding dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a plant branched-chain amino acid biosynthetic enzyme

10   selected from the group consisting of dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, and *leuD* subunit of 3-isopropylmalate dehydratase.

In another embodiment, the instant invention relates to a chimeric gene encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of

15   3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding a dihydroxyacid dehydratase, a branched chain amino acid amino-transferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase, operably linked to suitable regulatory sequences, wherein

20   expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a dihydroxyacid dehydratase, a branched

25   chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The

30   invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of

35   3-isopropylmalate dehydratase in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase; and

2

b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase in the transformed host cell; (c) optionally purifying the dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase expressed by the transformed host cell; (d) treating the dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase with a compound to be tested; and (e) comparing the activity of the dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase that has been treated with a test compound to the activity of an untreated dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE
DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

3

Figure 1 depicts the metabolic pathway leading to biosynthesis and utilization of the branched chain amino acids isoleucine, valine and leucine. The following abbreviations are used: ALS = acetolactate synthase; AHRI = acetohydroxyacid reductoisomerase; DHAD = dihydroxyacid dehydratase; BCAT = branched chain amino acid aminotransferase; IPMS = isporpoylmalate synthase; IPMD = 3-isopropylmalate dehydratase; and IPMDH = isopropylmalate deyhdrogenase.

Figure 2 depicts the amino acid sequence alignments between the dihydroxyacid dehydratase from corn clone cr1.pk0032.c4 (SEQ ID NO:2), soybean contig assembled from clones se3.pk0006.g4, and ses9c.pk001.o8 (SEQ ID NO:4), wheat clone wkm2c.pk005.c12 (SEQ ID NO:6), and *Saccharomyces cerevisiae* (NCBI gi Accession No. 1170543, SEQ ID NO:7). Amino acids which are conserved among all sequences are indicated with a plus sign (+) while those conserved only within the plant sequences are indicated by an asterisk (*).

Figure 3 depicts the amino acid sequence alignments between the branched chain amino acid transaminase from corn clone cc71se-b.pk0008.b5 (SEQ ID NO:9), corn clone cen6.pk0003.b5 (SEQ ID NO:11), corn clone cta1n.pk0070.e7 (SEQ ID NO:13), rice clone rls24.pk0025.f6 (SEQ ID NO:15), soybean clone ses8w.pk0032.e9 (SEQ ID NO:17), wheat clone wlm96.pk027.n2 (SEQ ID NO:19), and *Bacillus subtilis* (NCBI gi Accession No. 1706292, SEQ ID NO:20). Amino acids which are conserved among all sequences are indicated with a plus sign (+) while those conserved only within the plant sequences are indicated by an asterisk (*).

Figure 4 depicts the amino acid sequence alignments between the branched chain amino acid transaminase from rice clone rls72.pk0014.a3 (SEQ ID NO:22), soybean clone sre.pk0001.d1 (SEQ ID NO:24), soybean clone srr2c.pk003.d20 (SEQ ID NO:26), wheat clone wl1n.pk0123.c11 (SEQ ID NO:28), and *Methanococcus jannaschii* (NCBI gi Accession No. 124380, SEQ ID NO:29). Amino acids which are conserved among all sequences are indicated with a plus sign (+) while those conserved only within the plant sequences are indicated by an asterisk (*).

Figure 5 depicts the amino acid sequence alignments between the branched chain amino acid transaminase from corn clone cco1.pk0030.d2 (SEQ ID NO:31), wheat clone wkm1c.pk0004.c7 (SEQ ID NO:33), and *Escherichia coli* (NCBI gi Accession No. 1705437, SEQ ID NO:34). Amino acids which are conserved among all sequences are indicated with a plus sign (+) while those conserved only within the plant sequences are indicated by an asterisk (*).

Figure 6 depicts the amino acid sequence alignments between the *leuC* subunit of 3-isopropylmalate dehydratase from the corn contig assembled from clones cen1.pk0032.b1, cta1n.pk0077.c7, cen3n.pk0015.g3, cen3n.pk0060.h4, and cen3n.pk0121.h11 (SEQ ID NO:36), corn clone cr1n.pk0153.e9 (SEQ ID NO:38), rice clone rl0n.pk087.k16 (SEQ ID NO:40), soybean contig assembled from clones sdp3c.pk008.k13, and srm.pk0006.h5 (SEQ

ID NO:42), wheat clone wre1n.pk0045.b10 (SEQ ID NO:44), and *Rhizomucor pusillus* (NCBI gi Accession No. 1708799, SEQ ID NO:45). Amino acids which are conserved among all sequences are indicated with a plus sign (+) while those conserved only within the plant sequences are indicated by an asterisk (*).

Figure 7 depicts the amino acid sequence alignments between the *leuD* subunit of 3-isopropylmalate dehydratase from corn clone cr1n.pk0123.b7 (SEQ ID NO:47), rice clone rls12.pk0001.c2 (SEQ ID NO:49), soybean clone srr1c.pk003.c2 (SEQ ID NO:51), wheat clone wl1n.pk0048.a6 (SEQ ID NO:53); and *Lactococcus lactis* (NCBI gi Accession No. 400187, SEQ ID NO:54). Amino acids which are conserved among all sequences are indicated with a plus sign (+) while those conserved only within the plant sequences are indicated by an asterisk (*).

The following sequence descriptions and the Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence comprising the entire cDNA insert in clone cr1.pk0032.c4 encoding a full-length corn dihydroxyacid dehydratase.

SEQ ID NO:2 is the deduced amino acid sequence of a corn dihydroxyacid dehydratase derived from the nucleotide sequence of SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence comprising the contig formed from a portion of the cDNA insert in clone ses9c.pk001.o8 and the entire cDNA insert in clone se3.pk0006.g4 encoding a full-length soybean dihydroxyacid dehydratase.

SEQ ID NO:4 is the deduced amino acid sequence of a soybean dihydroxyacid dehydratase derived from the nucleotide sequence of SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence comprising a portion of the cDNA insert in clone wkm2c.pk005.c12 encoding the C-terminal 156 amino acid from a wheat dihydroxyacid dehydratase.

SEQ ID NO:6 is the deduced amino acid sequence of a fragment from a wheat dihydroxyacid dehydratase derived from the nucleotide sequence of SEQ ID NO:5.

SEQ ID NO:7 is the amino acid sequence of a *Saccharomyces cerevisiae* dihydroxyacid dehydratase (NCBI gi Accession No. 1170543).

SEQ ID NO:8 is the nucleotide sequence comprising a portion of the cDNA insert in clone cc71se-b.pk0008.b5 encoding a fragment of a corn branched chain amino acid transaminase.

SEQ ID NO:9 is the deduced amino acid sequence of a fragment of a corn branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:8.

SEQ ID NO:10 is the nucleotide sequence comprising the entire cDNA insert in clone cen6.pk0003.b5 encoding the C-terminal half of a corn branched chain amino acid transaminase.

5

SEQ ID NO:11 is the deduced amino acid sequence of the C-terminal half of a corn branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:10.

SEQ ID NO:12 is the nucleotide sequence comprising a portion of the cDNA insert in clone cta1n.pk0070.e7 encoding the C-terminal 110 amino acids from a corn branched chain amino acid transaminase.

SEQ ID NO:13 is the deduced amino acid sequence of the C-terminal 110 amino acids from a corn branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:12.

SEQ ID NO:14 is the nucleotide sequence comprising the entire cDNA insert in clone rls24.pk0025.f6 encoding the C-terminal 75% of a rice branched chain amino acid transaminase.

SEQ ID NO:15 is the deduced amino acid sequence of the C-terminal 75% of a rice branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:16.

SEQ ID NO:16 is the nucleotide sequence comprising the entire cDNA insert in clone ses8w.pk0032.e9 encoding the N-terminal 60% of a soybean branched chain amino acid transaminase.

SEQ ID NO:17 is the deduced amino acid sequence of the N-terminal 60% of a soybean branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:16.

SEQ ID NO:18 is the nucleotide sequence comprising the entire cDNA insert in clone wlm96.pk027.n2 encoding the C-terminal 93% of a wheat branched chain amino acid transaminase.

SEQ ID NO:19 is the deduced amino acid sequence of the C-terminal 93% of a wheat branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:18.

SEQ ID NO:20 is the amino acid sequence of a *Bacillus subtilis* branched chain amino acid transaminase (NCBI gi Accession No. 1176947).

SEQ ID NO:21 is the nucleotide sequence comprising the entire cDNA insert in clone rls72.pk0014.a3 encoding the C-terminal 82% of a rice branched chain amino acid transaminase.

SEQ ID NO:22 is the deduced amino acid sequence of the C-terminal 82% of a rice branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:21.

SEQ ID NO:23 is the nucleotide sequence comprising the entire cDNA insert in clone sre.pk0001.d1 encoding a full-length soybean branched chain amino acid transaminase.

6

SEQ ID NO:24 is the deduced amino acid sequence of a full-length soybean branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:23.

SEQ ID NO:25 is the nucleotide sequence comprising the entire cDNA insert in clone srr2c.pk003.d20 encoding a full length soybean branched chain amino acid transaminase.

SEQ ID NO:26 is the deduced amino acid sequence of a full length soybean branched chain amino acid transaminase derived from the nucleotide sequence of. SEQ ID NO:25.

SEQ ID NO:27 is the nucleotide sequence comprising the entire cDNA insert in clone wl1n.pk0123.c11 encoding approximately the C-terminal 80% of a wheat branched-chain amino acid transaminase.

SEQ ID NO:28 is the deduced amino acid sequence of approximately the C-terminal 80% of a wheat branched-chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:27.

SEQ ID NO:29 is the amino acid sequence of a *Methanococcus jannaschii* branched chain amino acid transaminase (NCBI gi Accession No. 3122287).

SEQ ID NO:30 is the nucleotide sequence comprising the entire cDNA insert in clone cco1.pk0030.d2 encoding a full-length corn branched chain amino acid transaminase.

SEQ ID NO:31 is the deduced amino acid sequence of a full-length corn branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:30.

SEQ ID NO:32 is the nucleotide sequence comprising the entire cDNA insert in clone wkm1c.pk0004.c7 encoding the C-terminal 80% of a wheat branched chain amino acid transaminase.

SEQ ID NO:33 is the deduced amino acid sequence of the C-terminal 80% of a wheat branched chain amino acid transaminase derived from the nucleotide sequence of SEQ ID NO:32.

SEQ ID NO:34 is the amino acid sequence of a *Escherichia coli* branched chain amino acid transaminase (NCBI gi Accession No. 124380).

SEQ ID NO:35 is the nucleotide sequence comprising the contig assembled from the cDNA insert in clones cen1.pk0032.b1, cta1n.pk0077.c7, cen3n.pk0015.g3, cen3n.pk0060.h4, and cen3n.pk0121.h11 encoding a nearly full-length corn *leuC* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:36is the deduced amino acid sequence of a nearly full-length corn *leuC* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:35.

SEQ ID NO:37 is the nucleotide sequence comprising the entire cDNA insert in clone cr1n.pk0153.e9 encoding a full-length corn *leuC* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:38 is the deduced amino acid sequence of a full-length corn *leuC* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:37.

SEQ ID NO:39 is the nucleotide sequence comprising a portion of the cDNA insert in clone rl0n.pk087.k16 encoding a fragment of a rice *leuC* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:40 is the deduced amino acid sequence of a fragment of a rice *leuC* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:39.

SEQ ID NO:41 is the nucleotide sequence comprising a contig assembled from the cDNA insert in clones sdp3c.pk008.k13, and srm.pk0006.h5 encoding the C-terminal half of a soybean *leuC* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:42 is the deduced amino acid sequence of the C-terminal half of a soybean *leuC* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:41.

SEQ ID NO:43 is the nucleotide sequence comprising a portion of the cDNA insert in clone wre1n.pk0045.b10 encoding a portion of a wheat *leuC* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:44 is the deduced amino acid sequence of a portion of a wheat *leuC* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:43.

SEQ ID NO:45 is the amino acid sequence of a *Rhizomucor pusillus leuC* subunit of 3-isopropylmalate dehydratase (NCBI gi Accession No. 1708799).

SEQ ID NO:46 is the nucleotide sequence comprising the entire cDNA insert in clone cr1n.pk0123.b7 encoding a full-length corn *leuD* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:47 is the deduced amino acid sequence of a full-length corn *leuD* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:46.

SEQ ID NO:48 is the nucleotide sequence comprising the entire cDNA insert in clone rls12.pk0001.c2 encoding a full-length rice *leuD* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:49 is the deduced amino acid sequence of a full-length rice *leuD* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:48.

SEQ ID NO:50 is the nucleotide sequence comprising the entire cDNA insert in clone srr1c.pk003.c2 encoding a full-length soybean *leuD* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:51 is the deduced amino acid sequence of a full-length soybean *leuD* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:50.

SEQ ID NO:52 is the nucleotide sequence comprising the entire cDNA insert in clone wl1n.pk0048.a6 encoding a full-length wheat *leuD* subunit of 3-isopropylmalate dehydratase.

SEQ ID NO:53 is the deduced amino acid sequence of a full-length wheat *leuD* subunit of 3-isopropylmalate dehydratase derived from the nucleotide sequence of SEQ ID NO:52.

SEQ ID NO:54 is the amino acid sequence of a *Lactococcus lactis leuD* subunit of
5   3-isopropylmalate dehydratase located in NCBI gi Accession No. 400187.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research 13*:3021-3030 (1985) and in the *Biochemical Journal 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The
10   symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

<u>DETAILED DESCRIPTION OF THE INVENTION</u>

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-
15   stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify
20   common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence.
25   "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional
30   properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression
35   of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not

9

effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein. Preferred substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are 80% identical to the coding sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are 90% identical to the coding sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are 95% identical to the coding sequence of the nucleic acid fragments reported herein.

A "substantial portion" of an amino acid or nucleotide sequence comprises an effective length of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to conduct correlation assessment and putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may

10

now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

5   "Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of

10   3-isopropylmalate dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase proteins as set forth in SEQ ID NOs:2, 4, 6, 9, 11, 13, 15, 17, 19, 22, 24, 26, 28, 31, 33, 36, 38, 40, 42, 44, 47, 49, 51, and 53. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to

15   design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

        "Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically

20   assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on

25   optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

        "Gene" refers to a nucleic acid fragment that expresses a specific protein, including

30   regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that

35   are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but

that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G.D. (1995) *Molecular Biotechnology 3*:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell 1*:671-680.

12

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Pat. No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

13

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an

5   amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J.J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal

10   peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of

15   methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Pat. No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well

20   known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several plant branched chain amino acid biosynthetic enzymes have been isolated and identified by comparison of random

25   plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1
Branched Chain Amino Acid Biosynthetic Enzymes

Enzyme	Clone	Plant
Dihydroxyacid dehydratase	cr1.pk0032.c4	corn
	Contig of: se3.pk0006.g4 ses9c.pk001.o8	soybean
	wkm2c.pk005.c12	wheat
Branched chain amino acid transferase	cc71se-b.pk0008.b5	corn
	cco1.pk0030.d2	corn
	cen6.pk0003.b5	corn
	cta1n.pk0070.e7	corn
	rls24.pk0025.f6	rice
	rls72.pk0014.a3	rice
	ses8w.pk0032.e9	soybean
	sre.pk0001.d1	soybean
	srr2c.pk003.d20	soybean
	wkm1c.pk0004.c7	wheat
	wl1n.pk0123.c11	wheat
	wlm96.pk027.n2	wheat
leuC subunit of 3-isopropylmalate dehydratase (large subunit)	Contig of: cen1.pk0032.b1 cta1n.pk0077.c7 cen3n.pk0015.g3 cen3n.pk0060.h4 cen3n.pk0121.h11	corn
	cr1n.pk0153.e9	corn
	rl0n.pk087.k16	rice
	Contig of: sdp3c.pk008.k13 srm.pk0006.h5	soybean
	wre1n.pk0045.b10	wheat
leuD subunit of 3-isopropylmalate dehydratase (small subunit)	cr1n.pk0123.b7	corn
	rls12.pk0001.c2	rice
	srr1c.pk003.c2	soybean
	wl1n.pk0048.a6	wheat

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid

hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other dihydroxyacid dehydratase, branched chain amino

5   acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid

10   sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant

15   sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding

20   homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based

25   upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *PNAS USA 85*:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3'

30   or 5' cDNA fragments can be isolated (Ohara et al., (1989) *PNAS USA 86*:5673; Loh et al., (1989) *Science 243*:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M.A. and Martin, G.R., (1989) *Techniques 1*:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates

35   immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then

16

be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R.A. (1984) *Adv. Immunol. 36*:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of branched chain amino acids in those cells. Overexpression of dihydroxyacid dehydratase may produce a greater accumulation of 2-oxo-3-methylvalerate and 2-oxoisovalerate. These two products are substrates for branched chain amino acid aminotransferase which will become the limiting factor to producing higher amounts of leucine, valine, and isoleucine. Overexpression of the 3-isopropylmalate dehydratase subunits will lead to a greater accumulation of leucine precursors, and ultimately to a greater amount of leucine in the cell.

Overexpression of the dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J. 4*:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant branched chain amino acid biosynthetic enzymes to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode dihydroxyacid dehydratase, a

17

branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J.J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys.100*:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant amino acid biosynthetic enzymes can be constructed by linking a gene or gene fragment encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase in situ in cells or in vitro in cell extracts. Preferred heterologous host cells for production of the instant dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase. This chimeric gene could then be introduced into

appropriate microorganisms via transformation to provide high level expression of the encoded biosynthetic enzyme. An example of a vector for high level expression of the instant dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase in a bacterial host is provided (Example 9).

Additionally, the instant dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, or *leuD* subunit of 3-isopropylmalate dehydratase can be used as a targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate dehydratase, and the *leuD* subunit of 3-isopropylmalate dehydratase described herein catalyze various steps in the biosynthesis and utilization of branched chain amino acids. Accordingly, inhibition of the activity of one or more of the enzymes described herein could lead to inhibition plant growth. Thus, the instant dihydroxyacid dehydratase, branched chain amino acid aminotransferase, *leuC* subunit of 3-isopropylmalate dehydratase, and *leuD* subunit of 3-isopropylmalate dehydratase could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et at., (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter 4(1)*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and

19

other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet. 7*:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research 5*:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med. 114(2)*:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics 16*:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science 241*:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res. 18*:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics 7*:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res. 17*:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA 86*:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA 92*:8149; Bensen et al., (1995) *Plant Cell 7*:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the dihydroxyacid dehydratase, the branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate

dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-

5   anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a dihydroxyacid dehydratase, a branched chain amino acid aminotransferase, a *leuC* subunit of 3-isopropylmalate dehydratase, or a *leuD* subunit of 3-isopropylmalate dehydratase can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the dihydroxyacid dehydratase, the

10   branched chain amino acid aminotransferase, the *leuC* subunit of 3-isopropylmalate dehydratase, or the *leuD* subunit of 3-isopropylmalate dehydratase gene product.

<div align="center">EXAMPLES</div>

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be

15   understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

20   <div align="center">EXAMPLE 1</div>

<div align="center">Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones</div>

cDNA libraries representing mRNAs from various corn, rice, soybean, and wheat tissues were prepared. The characteristics of the libraries are described below.

25   <div align="center">TABLE 2</div>

<div align="center">cDNA Libraries from Corn, Rice, Soybean, and Wheat</div>

Library	Tissue	Clone
cc71se-b	Corn Callus Type II Tissue, Somatic Embryo Formed	cc71se-b.pk0008.b5
cco1	Corn Cob of 67 Day Old Plants Grown in Green House	cco1.pk0030.d2
cen1	Corn Endosperm 10 to 11 Days After Pollination	cen1.pk0032.b1
cen3n	Corn Endosperm 20 Days After Pollination*	cen3n.pk0015.g3 cen3n.pk0060.h4 cen3n.pk0121.h11
cen6	Corn Developing Embryo 40 Days After Pollination	cen6.pk0003.b5
crl	Corn Root From 7 Day Old Seedlings	crl.pk0032.c4
crln	Corn Root From 7 Day Old Seedlings*	crln.pk0123.b7 crln.pk0153.e9
ctaln	Corn Tassel*	ctaln.pk0070.e7 ctaln.pk0077.c7
rl0n	Rice 15 Day Old Leaf*	rl0n.pk087.k16

<div align="center">21</div>

rls12	Rice Leaf 15 Days After Germination, 12 hours after infection of strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO); Susceptible	rls12.pk0001.c2
rls24	Rice Leaf 15 Days After Germination, 24 Hours After Infection of Strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO); Susceptible	rls24.pk0025.f6
rls72	Rice Leaf 15 Days After Germination, 72 Hours After Infection of Strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO); Susceptible	rls72.pk0014.a3
sdp3c	Soybean Developing Pods (8-9 mm)	sdp3c.pk008.k13
se3	Soybean Embryo, 17 Days After Flowering	se3.pk0006.g4
ses8w	Soybean Mature Embryo 8 Weeks After Subculture	ses8w.pk0032.e9
ses9c	Soybean Embryogenic Suspension	ses9c.pk001.o8
sre	Soybean Root Elongation Zone 4 to 5 Days After Germination	sre.pk0001.d1
srm	Soybean Root Meristem	srm.pk0006.h5
srr1c	Soybean 8-Day-Old Root	srr1c.pk003.c2
srr2c	Soybean 8-Day-Old Root	srr2c.pk003.d20
wkm1c	Wheat Kernel Malted 55 Hours at 22 Degrees Celsius	wkm1c.pk0004.c7
wkm2c	Wheat Kernel Malted 175 Hours at 4 Degrees Celsius	wkm2c.pk005.c12
wl1n	Wheat Leaf From 7 Day Old Seedling*	wl1n.pk0048.a6 wl1n.pk0123.c11
wlm96	Wheat Seedlings 96 Hours After Inoculation With *Erysiphe graminis f. sp tritici*	wlm96.pk027.n2
wre1n	Wheat Root From 7 Day Old Etiolated Seedling*	wre1n.pk0045.b10

*These libraries were normalized essentially as described in U.S. Pat. No. 5,482,845, incorporated herein by reference

cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the

5   manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using

10   primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science 252*:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

22

EXAMPLE 2
Identification of cDNA Clones

ESTs encoding branched chain biosynthetic enzymes were identified by conducting
BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.*
215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences
contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS
translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data
Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and
DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity
to all publicly available DNA sequences contained in the "nr" database using the BLASTN
algorithm provided by the National Center for Biotechnology Information (NCBI). The
DNA sequences were translated in all reading frames and compared for similarity to all
publicly available protein sequences contained in the "nr" database using the BLASTX
algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics 3*:266-272) provided by the
NCBI. For convenience, the P-value (probability) of observing a match of a cDNA
sequence to a sequence contained in the searched databases merely by chance as calculated
by BLAST are reported herein as "pLog" values, which represent the negative of the
logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the
likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3
Characterization of cDNA Clones Encoding Dihydroxyacid Dehydratase

The BLASTX search using the nucleotide sequences from clones se3.pk0006.g4,
ses4d.pk0014.h5, ssl.pk0021.g5, sre.pk0012.h2, ses8w.pk0010.e5, se4.03a05,
ssm.pk0004.h12, ssm.pk0033.d5, cr1.pk0032.c4, cen1.pk0015.f5, cen3n.pk0071.h6,
cen3n.pk0031.g9 and wre1n.pk0007.a4 revealed similarity of the proteins encoded by the
cDNAs to dihydroxyacid dehydratases from several organisms, including *Lactococcus lactis*
(SWISS-PROT Accession No. Q02139, NCBI gi Accession No. 400054) and
Saccharomyces cerevisiae (SWISS-PROT Accession No. P39522, NCBI gi Accession
No. 1170543). The BLAST results for each of these ESTs are shown in Table 3:

TABLE 3
BLAST Results for Clones Encoding Polypeptides Homologous
to Dihydroxyacid Dehydratases

| | BLAST pLog Score | |
Clone	Q02139 (*Lactococcus lactis*)	P39522 (*Saccharomyces cerevisiae*)
cen1.pk0015.f5	41.03	21.43
cen3n.pk0031.g9	38.89	40.35
cen3n.pk0071.h6	32.62	34.55

cr1.pk0032.c4	5.64	8.33
se3.pk0006.g4	26.64	37.80
se4.03a05	12.41	19.03
ses4d.pk0014.h5	4.89	na*
ses8w.pk0010.e5	37.34	33.74
sre.pk0012.h2	41.54	37.96
ssl.pk0021.g5	38.66	33.92
ssm.pk0004.h12	5.74	10.82
ssm.pk0033.d5	23.30	28.59
wre1n.pk0007.a4	6.39	7.29

na* = no data available

The sequence of the entire cDNA insert in clone cr1.pk0032.c4 was determined and is shown in SEQ ID NO:1. This sequence encodes a full-length dihydroxyacid dehydratase, and includes the sequences from clones cen1.pk0015.f5, cen3n.pk0031.g9, and cen3n.pk0071.h6. The deduced amino acid sequence of this cDNA is shown in SEQ ID NO:2. The amino acid sequence set forth in SEQ ID NO:2 was evaluated by BLASTP, yielding a pLog value larger than 254 versus the *Saccharomyces cerevisiae* sequence.

A sequence encoding a full-length dihydroxyacid dehydratase was formed by assembling a contig with the nucleotides of the entire cDNA insert in clone se3.pk0006.g4 and a portion of the cDNA insert from clone ses9c.pk001.o8. This sequence includes the nucleotide sequences from clones se4.03a05, ses4d.pk0014.h5, ses8w.pk0010.e5, sre.pk0012.h2, ssl.pk0021.g5, and ssm.pk0004.h12. The nucleotide sequence of this contig is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:4. The amino acid sequence set forth in SEQ ID NO:4 was evaluated by BLASTP, yielding a pLog value larger than 254 versus the *Saccharomyces cerevisiae* sequence.

The sequence of a portion of the cDNA insert from clone wkm2c.pk005.c12 was determined and found to contain the sequence from clone wre1n.pk0007.a4. The BLASTX search using the nucleotide sequence from clone wkm2c.pk005.c12 resulted in a pLog value of 49 against the *Saccharomyces cerevisiae* dihydroxyacid dehydratase and a pLog value of 47 against the *Lactococcus lactis* sequence. The sequence for clone wkm2c.pk005.c12 is shown in SEQ ID NO:5; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:6.

Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4, and 6 and the *Saccharomyces cerevisiae* sequence (SEQ ID NO:7). The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, and 6 and the *Saccharomyces cerevisiae* dihydroxyacid dehydratase sequence.

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of
cDNA Clones Encoding Polypeptides Homologous
to Dihydroxyacid Dehydratase

Clone	SEQ ID NO.	Percent Identity to NCBI gi Accession No. 1170543 (*Saccharomyces cerevisiae*)
cr1.pk0032.c4	2	58.5
Contig of: se3.pk0006.g4 ses9c.pk001.o8	4	58.8
wkm2c.pk005.c12	6	55.8

Sequence alignments and percent identity calculations were performed using the
Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc.,
Madison, WI). Multiple alignment of the sequences was performed using the Clustal method
of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS. 5*:151-153) using the default
parameters (GAP PENALTY=10, GAP LENGHT PENALTY=10). Sequence percent
identity calculations were performed by the Jotun Hein method (Hein. J. J. (1990) *Meth. Enz.
183*:626-645). Sequence alignments and BLAST scores and probabilities indicate that the
instant nucleic acid fragments encode entire corn and soybean dihydroxyacid dehydratase
and a portion of a wheat dihydroxyacid dehydratase. These sequences represent the first
plant sequences encoding a dihydroxyacid dehydratase enzyme.

EXAMPLE 4

Characterization of cDNA Clones Encoding
Branched-Chain Amino Acid Aminotransferases

The BLASTX search using the nucleotide sequences of clones ses8w.pk0032.e9,
cco1.pk0030.d2 and cen6.pk0003.b5 revealed similarity of the proteins encoded by the
cDNAs to putative branched chain amino acid aminotransferases from *Bacillus subtilis*
(SWISS-PROT Accession No. P39576, NCBI gi Accession No. 1176947) and
Mycobacterium tuberculosis (SWISS-PROT Accession No. Q10399, NCBI gi Accession
No. 1708468). A further BLASTX search using the nucleotide sequences from clones
cc71se-b.pk0008.b5, cta1n.pk0070.e7, rls24.pk0025.f6, and wlm96.pk027.n2 also revealed
similarity of the proteins encoded by the cDNAs to putative branched chain amino acid
aminotransferases from *Bacillus subtilis* and *Mycobacterium tuberculosis*. The BLASTX
search using the nucleotide sequences from clones rls72.pk0014.a3, sre.pk0001.d1,
srr2c.pk003.d20, and wl1n.pk0123.c11 revealed similarity of the proteins encoded by the
cDNAs to putative branched chain amino acid aminotransferase from *Methanococcus
jannaschii* (NCBI gi Accession No. 3122287). The BLASTX search using the nucleotide

sequences from clones cco1.pk0030.d2, and wkm1c.pk0004.c7 revealed similarity of the proteins encoded by the cDNAs to branched chain amino acid aminotransferase from *Escherichia coli* (NCBI gi Accession No. 124380). The BLASTX results for each of these sequences are shown in Table 5:

TABLE 5
BLASTX Results for Clones Encoding Polypeptides Homologous to Branched-Chain Amino Acid Aminotransferases

Clone	Accession No.	BLAST pLog Score	Accession No.	BLAST pLog Score
cc71se-b.pk0008.b5	1176947	16.22	1708468	18.10
cen6.pk0003.b5	1176947	50.40	1708468	34.30
cta1n.pk0070.e7	1176947	12.00	1708468	10.50
rls24.pk0025.f6	1176947	66.00	1708468	56.70
ses8w.pk0032.e9	1176947	41.70	1708468	39.70
wlm96.pk027.n2	1176947	88.40	1708468	67.00
rls72.pk0014.a3	3122287	9.00		
sre.pk0001.d1	3122287	24.00		
srr2c.pk003.d20	3122287	21.70		
wl1n.pk0123.c11	3122287	9.40		
cco1.pk0030.d2	124380	26.70		
wkm1c.pk0004.c7	124380	19.05		

The sequence of a portion of the cDNA insert in clone cc71se-b.pk0008.b5 was determined and is shown in SEQ ID NO:8; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:9. The sequence of the entire cDNA insert in clone cen6.pk0003.b5 was determined and is shown in SEQ ID NO:10; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:11. The sequence of a portion of the cDNA insert in clone cta1n.pk0070.e7 was determined and is shown in SEQ ID NO:12; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:13. The sequence of the entire cDNA insert in clone rls24.pk0025.f6 was determined and is shown in SEQ ID NO:14; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:15. The sequence of the entire cDNA insert in clone ses8w.pk0032.e9 was determined and is shown in SEQ ID NO:16; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:17. The sequence of the entire cDNA insert in clone wlm96.pk027.n2 was determined and is shown in SEQ ID NO:18; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:19. The amino acid sequences set forth in SEQ ID NO:11, SEQ ID NO:15, SEQ ID NO:17, and SEQ ID NO:19 were evaluated by BLASTP revealing similarity to the branched-chain amino acid aminotransferase from *Bacillus subtilis* (NCBI gi Accession No. 1176947); the results of

26

these analyses are shown in Table 6. Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:9, 11, 13, 15, 17, and 19 and the *Bacillus subtilis* sequence (SEQ ID NO:20).

The sequence of the entire cDNA insert in clone rls72.pk0014.a3 was determined and is shown in SEQ ID NO:21; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:22. The sequence of the entire cDNA insert in clone sre.pk0001.d1 was determined and is shown in SEQ ID NO:23; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:24. The sequence of the entire cDNA insert in clone srr2c.pk003.d20 was determined and is shown in SEQ ID NO:25; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:26. The sequence of the entire cDNA insert in clone wl1n.pk0123.c11 was determined and is shown in SEQ ID NO:27; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:28. The amino acid sequences set forth in SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, and SEQ ID NO:28, were evaluated by BLASTP revealing similarity to the branched-chain amino acid aminotransferase from *Methanococcus jannaschii* (NCBI gi Accession No. 3122287), the results of these analyses are shown in Table 6. Figure 4 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:22, 24, 26, and 28 and the *Methanococcus jannaschii* sequence (SEQ ID NO:29).

The sequence of the entire cDNA insert in clone cco1.pk0030.d2 was determined and is shown in SEQ ID NO:30; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:31. The sequence of the entire cDNA insert in clone wkm1c.pk0004.c7 was determined and is shown in SEQ ID NO:32; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:33. The amino acid sequences set forth in SEQ ID NO:31, and SEQ ID NO:33 were evaluated by BLASTP revealing similarity to the branched-chain amino acid aminotransferase from *Escherichia coli* (NCBI gi Accession No. 124380), the results of these analyses are shown in Table 6. Figure 5 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:31, and 33 and the *Escherichia coli* sequence (SEQ ID NO:34).

TABLE 6

BLASTP Results for Clones Encoding Polypeptides Homologous
to Branched-Chain Amino Acid Aminotransferases

Clone	Accession No.	BLASTP pLog Score
cen6.pk0003.b5	1176947	45.00
rls24.pk0025.f6	1176947	63.00
ses8w.pk0032.e9	1176947	35.00
wlm96.pk027.n2	1176947	77.22
rls72.pk0014.a3	3122287	11.70
sre.pk0001.d1	3122287	24.22

27

srr2c.pk003.d20	3122287	22.52
wlln.pk0123.c11	3122287	12.40
cco1.pk0030.d2	124380	31.00
wkm1c.pk0004.c7	124380	21.00

The data in Table 7 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:9, 11, 13, 15, 17, and 19 and the *Bacillus subtilis* sequence; a percent identity of the amino acid sequences set forth in SEQ ID NOs:22, 24, 5   26, and 28 and the *Methanococcus jannaschii* sequence; and a percent identity of the amino acid sequences set forth in SEQ ID NOs:31, and 33 and the *Escherichia coli* sequence.

TABLE 7

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences
10   of cDNA Clones Encoding Polypeptides Homologous to
Branched-Chain Amino Acid Aminotransferases

Clone	SEQ ID NO.	Accession No.	Percent Identity
cc71se-b.pk0008.b5	9	1176947	11.3
cen6.pk0003.b5	11	1176947	45.2
cta1n.pk0070.e7	13	1176947	41.1
rls24.pk0025.f6	15	1176947	40.6
ses8w.pk0032.e9	17	1176947	12.2
wlm96.pk027.n2	19	1176947	44.7
rls72.pk0014.a3	22	3122287	25.5
sre.pk0001.d1	24	3122287	30.7
srr2c.pk003.d20	26	3122287	26.3
wlln.pk0123.c11	28	3122287	30.6
cco1.pk0030.d2	31	124380	32.7
wkm1c.pk0004.c7	33	124380	30.8

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., 15   Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS. 5*:151-153) using the default parameters (GAP PENALTY=10, GAP LENGHT PENALTY=10). Sequence percent identity calculations were performed by the Jotun Hein method (Hein. J. J. (1990) *Meth. Enz. 183*:626-645).

20   Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode three distinct classes of branched-chain amino acid aminotransferases. Fragments from corn, rice, soybean, and wheat are similar to the *Bacillus subtilis* putative branched-chain amino acid aminotransferase, full-length soybean, and

fragments from rice, and wheat similar to the *Methanococcus jannaschii* enzyme, and a full-length corn, and most of a wheat branched chain amino acid aminotransferase similar to the *Escherichia coli* enzyme. These sequences represent the first plant sequences encoding branched-chain amino acid aminotransferase enzymes.

EXAMPLE 5

Characterization of cDNA Clones Encoding
Large Subunit of 3-Isopropylmalate Dehydratase

The BLASTX search using the nucleotide sequences from clones srm.pk0006.h5, ceb3.pk0001.b5, cen1.pk0032.b1, cen3n.pk0015.g3 and cen3n.pk0060.h4 revealed similarity of the proteins encoded by the cDNAs to the large subunit of 3-isopropylmalate dehydratase, which is encoded by the gene designated *leuC* in *E. coli* and *Salmonella typhumurium*. Similarity of the proteins is greatest to enzymes encoded by GenBank Accession No. U67499 and GenBank Accession No. U67543 from *Methanococcus jannaschii* (NCBI gi Accession No. 3219823, and 3122347, respectively). Both of these GenBank Accessions show strong similarity to the *leuC* subunit of 3-isopropylmalate dehydratase, but have been mis-labelled as aconitase and homoaconitase. Neither aconitase nor homoaconitase are expected to be present in *Methanococcus jannaschii* (see Selkov et al. (1997) *Gene 197*:GC11-GC26; this paper can be accessed via the World Wide Web at http://www.cme.msu.edu/wit/Doc/mj_recon.html). The BLAST results for each of these ESTs are shown in Table 8:

TABLE 8
BLAST Results for Clones Encoding Polypeptides Homologous to the
leuC Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

Clone	BLAST pLog Score	
	U67499	U67543
srm.pk0006.h5	6.92	8.80
ceb3.pk0001.b5	52.08	23.68
cen1.pk0032.b1	44.89	43.01
cen3n.pk0015.g3	12.24	11.92
cen3n.pk0060.h4	14.17	14.74

The BLASTX search using the nucleotide sequences from the contig assembled of clones cen1.pk0032.b1, cta1n.pk0077.c7, cen3n.pk0015.g3, cen3n.pk0060.h4, and cen3n.pk0121.h11 (the clone ceb3.pk0001.b5 is included in the contig), the nucleotide sequences from clones cr1n.pk0153.e9, rl0n.pk087.k16, and wre1n.pk0045.b10, and the nucleotide sequences from the contig assembled of clones sdp3c.pk008.k13 and srm.pk0006.h5 revealed similarity of the same proteins. The BLAST results for each of these sequences are shown in Table 9:

TABLE 9

BLAST Results for Clones Encoding Polypeptides Homologous to the
leuC Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

Clone	BLASTX pLog Score	
	NCBI gi Accession No. 3219823	NCBI gi Accession No. 3122347

Contig of:
    cen1.pk0032.b1
    cta1n.pk0077.c7
    cen3n.pk0015.g3
    cen3n.pk0060.h4
    cen3n.pk0121.h11
	91.22	66.30

crln.pk0153.e9	94.22	69.22
rl0n.pk087.k16	44.00	27.40

Contig of:
    sdp3c.pk008.k13
    srm.pk0006.h5
	28.15	25.40

wreln.pk0045.b10	12.00	11.52

The sequence of the contig assembled of the cDNA insert in clones cen1.pk0032.b1, cta1n.pk0077.c7, cen3n.pk0015.g3, cen3n.pk0060.h4, and cen3n.pk0121.h11 was determined and is shown in SEQ ID NO:35; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:36. The amino acid sequence set forth in SEQ ID NO:36 was evaluated by BLASTP, yielding a pLog value of 93.00 versus the *Methanococcus jannaschii* sequence (NCBI gi Accession No. 3219823). The sequence of the entire cDNA insert in clone crln.pk0153.e9 was determined and is shown in SEQ ID NO:37; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:38. The amino acid sequence set forth in SEQ ID NO:38 was evaluated by BLASTP, yielding a pLog value of 91.00 versus the *Methanococcus jannaschii* sequence (NCBI gi Accession No. 3219823). The sequence of a portion of the cDNA insert in clone rl0n.pk087.k16 was determined and is shown in SEQ ID NO:39; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:40. The sequence of the contig assembled of the cDNA insert in clones sdp3c.pk008.k13 and srm.pk0006.h5 was determined and is shown in SEQ ID NO:41; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:42. The sequence of a portion of the cDNA insert in clone wreln.pk0045.b10 was determined and is shown in SEQ ID NO:43; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:44.

Figure 6 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:36, 38, 40, 42, and 44 and the *Methanococcus jannaschii* sequence (SEQ ID NO:45). The data in Table 10 represents a calculation of the percent identity of the amino acid

sequences set forth in SEQ ID NOs:36, 38, 40, 42, and 44 and the *Methanococcus jannaschii* sequence.

TABLE 10

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of
cDNA Clones Encoding Polypeptides Homologous
to the *leuC* Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

Clone	SEQ ID NO.	Percent Identity to NCBI gi Accession No. 3219823
Contig of: cen1.pk0032.b1 ctaln.pk0077.c7 cen3n.pk0015.g3 cen3n.pk0060.h4 cen3n.pk0121.h11	36	45.8
cr1n.pk0153.e9	38	45.6
rl0n.pk087.k16	40	55.5
Contig of: sdp3c.pk008.k13 srm.pk0006.h5	42	15.4
wre1n.pk0045.b10	44	16.5

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS. 5*:151-153) using the default parameters (GAP PENALTY=10, GAP LENGHT PENALTY=10). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz. 183*:626-645).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode two distinct entire corn the *leuC* subunit of 3-isopropylmalate dehydratases, a central fragment of a rice *leuC* subunit of 3-isopropylmalate dehydratase, and the C-terminal half of a soybean, and a wheat *leuC* subunit of 3-isopropylmalate dehydratase. These sequences represent the first plant sequences encoding *leuC* subunit of 3-isopropylmalate dehydratase enzyme.

EXAMPLE 6

Characterization of cDNA Clones Encoding the
leuD Subunit of 3-Isopropylmalate Dehydratase

The BLASTX search using the nucleotide sequences from EST clones cr1n.pk0123.b7 and rls12.pk0001.c2 revealed similarity of the proteins encoded by the cDNAs to the small subunit of 3-isopropylmalate dehydratase enzymes from several organisms, including *Methanococcus jannaschii* (GenBank Accession No. U67568) and *Lactococcus lactis*

(SWISS-PROT Accession No. Q02144). This subunit is encoded by the gene designated *leuD* in *E. coli* and *Salmonella typhimurium*. The BLAST results for each of these ESTs are shown in Table 11:

TABLE 11

BLAST Results for Clones Encoding Polypeptides Homologous to the
leuD Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

	BLAST pLog Score	
Clone	U67568	Q02144
cr1n.pk0123.b7	3.70	1.80
rls12.pk0001.c2	5.14	3.96

The BLASTX search using the nucleotide sequences of the entire insert from clones cr1n.pk0123.b7, rls12.pk0001.c2, srr1c.pk003.c2, and wl1n.pk0048.a6 also revealed similarity of the proteins encoded by the cDNAs to the small subunit of 3-isopropylmalate dehydratase enzymes from several organisms, including *Methanococcus jannaschii* (NCBI gi Accession No. 3122345) and *Lactococcus lactis* (NCBI gi Accession No. 400187). The BLAST results for each of these sequences are shown in Table 12:

TABLE 12

BLAST Results for Clones Encoding Polypeptides Homologous
to the *leuD* Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

	BLAST pLog Score	
Clone	NCBI gi Accession No. 3122345	NCBI gi Accession No. 400187
cr1n.pk0123.b7	16.70	12.30
rls12.pk0001.c2	16.40	11.30
srr1c.pk003.c2	18.70	11.52
wl1n.pk0048.a6	17.05	12.22

The sequence of the entire cDNA insert in clone cr1n.pk0123.b7 was determined and is shown in SEQ ID NO:46; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:47. The sequence of the entire cDNA insert in clone rls12.pk0001.c2 was determined and is shown in SEQ ID NO:48; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:49. The sequence of the entire cDNA insert in clone srr1c.pk003.c2 was determined and is shown in SEQ ID NO:50; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:51. The sequence of the entire cDNA insert in clone wl1n.pk0048.a6 was determined and is shown in SEQ ID NO:52 the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:53. The amino acid sequences set forth in SEQ ID NOs:47, 49, 51, and 53 were evaluated by BLASTP, yielding the pLog values

shown in Table 13 versus the *Methanococcus jannaschii* (NCBI gi Accession No. 3122345)
and *Lactococcus lactis* (NCBI gi Accession No. 400187) sequences.

TABLE 13

BLASTP Results for Clones Encoding Polypeptides Homologous
to the *leuD* Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

| | BLASTP pLog Score | |
Clone	NCBI gi Accession No. 3122345	NCBI gi Accession No. 400187
crln.pk0123.b7	16.70	12.30
rls12.pk0001.c2	16.40	11.30
srr1c.pk003.c2	18.70	11.52
wlln.pk0048.a6	17.05	12.22

Figure 7 presents an alignment of the amino acid sequences set forth in SEQ ID
NOs:47, 49, 51, and 53 and the *Lactococcus lactis* sequence (SEQ ID NO:54). The data in
Table 14 represents a calculation of the percent identity of the amino acid sequences set forth
in SEQ ID NOs:47, 49, 51, and 53 and the *Lactococcus lactis* sequence.

TABLE 14

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of
cDNA Clones Encoding Polypeptides Homologous
to the *leuD* Protein Subunit of 3-Isopropylmalate Dehydratase Enzymes

Clone	SEQ ID NO.	Percent Identity to 400187
crln.pk0123.b7	47	35.8
rls12.pk0001.c2	49	34.5
srr1c.pk003.c2	51	38.0
wlln.pk0048.a6	53	36.0

Sequence alignments and percent identity calculations were performed using the
Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc.,
Madison, WI). Multiple alignment of the sequences was performed using the Clustal method
of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS*. *5*:151-153) using the default
parameters (GAP PENALTY=10, GAP LENGHT PENALTY=10). Sequence percent
identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.*
183:626-645).

Sequence alignments and BLAST scores and probabilities indicate that the instant
nucleic acid fragments encode entire corn, rice, soybean, and wheat *leuD* protein subunit of
3-isopropylmalate dehydratase enzymes. These sequences represent the first plant sequences
encoding *leuD* protein subunit of 3-isopropylmalate dehydratase.

EXAMPLE 7
Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding a branched-chain biosynthetic enzyme in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a branched-chain biosynthetic enzyme, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236)

which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the gluphosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

35

EXAMPLE 8
Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from
5   the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant branched-chain amino acid biosynthetic enzymes in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction
10   endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment
15   when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding a branched-chain amino acid biosynthetic enzyme. To induce somatic
20   embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are
25   maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

30   Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Kline et al. (1987) *Nature* (London) 327:70, U.S. Patent No. 4,945,050). A Du Pont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a
35   chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression

cassette comprising the phaseolin 5' region, the fragment encoding the branched-chain amino acid enzyme and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL CaCl$_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

<u>EXAMPLE 9</u>

<u>Expression of Chimeric Genes in Microbial Cells</u>

The cDNAs encoding the instant branched-chain amino acid biosynthetic enzyme can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted

37

to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve

5 GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase

10 (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent

15 cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the branched-chain amino acid biosynthetic enzyme are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct

20 orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by

25 centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by

30 SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

<div align="center">EXAMPLE 10</div>

<div align="center">Evaluating Compounds for Their Ability to Inhibit the Activity</div>

<div align="center">of a Plant Branched Chain Amino Acid Biosynthetic Enzyme</div>

35 The branched-chain amino acid biosynthetic enzymes described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 9, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in

<div align="center">38</div>

suitably infected organisms or cell lines. The instant branched-chain amino acid biosynthetic enzymes may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin

5   ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("(His)$_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the

10   fusion protein and the enzyme.

Purification of the instant branched-chain amino acid biosynthetic enzyme, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate

15   precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the branched-chain amino acid biosynthetic enzymes are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the

20   expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, a branched-chain amino acid biosynthetic enzyme may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a (His)$_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by

25   linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated

30   above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic

35   activation of the branched-chain amino acid biosynthetic enzymes disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. Examples of assays for these enzymes can be found in *Methods in Enzymology* (1970) Vol. XVII, Part A, (Tabor and Tabor eds.) Academic Press, New York.

39

Assays for branched-chain amino acid transferase are presented by Jenkins et al., pp 802-807, Aki et al., pp 807-811, and Aki et al., pp 811-814 of the above volume. An assay for dihyroxyacid dehydratase is presented by Kiritani et al., pp 755-764. Assays for isopropylmalate dehydratase (which is refered to as isopropylmalate isomerase) are

5   presented by Gross, pp 786-790, and by Cho-Chung et al., pp 782-785.